Predictive Identification of Co-formers in Co-amorphous Systems

Luke I. Chambers[1], Holger Grohganz[2], Henrik Palmelund[2], Korbinian Löbmann[2], Thomas Rades[2], Osama M. Musa[3], Jonathan W. Steed[1*]

1) Durham University, Department of Chemistry, Lower Mountjoy, Stockton Road, Durham, DH1 3LE, UK.
2) Department of Pharmacy, University of Copenhagen, Copenhagen, Denmark.
3) Ashland LLC, 1005 Route 202/206, Bridgewater, NJ 08807, USA.

*Corresponding author. Email address: jon.steed@durham.ac.uk

Abstract

This work aims to understand the properties of co-formers that form co-amorphous pharmaceutical materials and to predict co-amorphous system formation. A partial least square – discriminant analysis (PLS-DA) was performed using known co-amorphous systems described by 36 variables based on the properties of the co-former and the binding energy of the system. The PLS-DA investigated the propensity to form co-amorphous material of the active pharmaceutical ingredients: mebendazole, carvedilol, indomethacin, simvastatin, carbamazepine and furosemide in combination with 20 amino acid co-formers. The variables that were found to favour the propensity to form co-amorphous systems appear to be a

relatively large value for average molecular weight and the sum of the difference between hydrogen bond donors and hydrogen bond acceptors for both components, and a relatively small or negative value for excess enthalpy of mixing, excess enthalpy of hydrogen bonding and the difference in the Hansen parameter for hydrogen bonding of the coformer and the active pharmaceutical ingredient (API). To test the predictive power of this model, 29 potential co-formers were used to form either co-amorphous or crystalline two-component materials with mebendazole. Of these 29 two-component systems, the co-amorphous nature of a total of 26 materials was correctly predicted by the model, giving a predictive hit rate of 90 %.

Keywords

Co-amorphous

Partial least squares discriminant analysis

Amino acids

Multi-variate analysis

Molecular descriptors

1. Introduction

A large proportion of newly discovered active pharmaceutical ingredients (APIs) display poor solubility in the gastrointestinal fluids, which is likely to decrease their bioavailability (Di et al., 2012; Kalepu and Nekkanti, 2015; Khadka et al., 2014; Savjani et al., 2012). To

improve the aqueous solubility of APIs, different formulation methods have been designed including amorphous forms, which have no long-range crystallographic order and higher internal energy compared with their respective crystalline forms (Berry and Steed, 2017; Healy et al., 2017; Khodadadi and Meesters, 2018; Williams et al., 2013). However, pure amorphous APIs are often physically unstable and can crystallise as a result of increased molecular mobility, especially when stored above their glass transition temperature or in humid environments (Kissi et al., 2018; Rams-Baron et al., 2018; Sun et al., 2012). Methods to improve the stability of amorphous APIs include the formation of amorphous solid dispersions and co-amorphous (COAM) materials (Karagianni et al., 2018; Ma and Williams, 2019; Van Den Mooter, 2012; Wu et al., 2018).

Amorphous solid dispersions are formed by (molecularly) dispersing an API in a (usually amorphous) polymer such as polyvinylpyrrolidone and cellulose based polymers, which act as an inactive stabilizer (Chavan et al., 2019; Nielsen et al., 2015; Vasconcelos et al., 2016). Stabilization (even above the solubility limit of the API in the polymer) is caused by the polymer increasing the glass transition temperature and forming intermolecular interactions, which in turn result in reduced molecular mobility (Baghel et al., 2016; Frank and Matzger, 2018; Medarević et al., 2019). The main challenges with using amorphous solid dispersions are their often high hygroscopicity (causing increased molecular mobility of the API), and the usually large mass ratios of polymer to API (causing downstream formulation problems when high API dosages are required) (Marsac et al., 2008; Rumondor et al., 2009; Tian et al., 2015).

COAM systems are formed by mixing an API with a low molecular weight compound called a co-former, which is usually inactive but could also be another API (Gao et al., 2013;

Newman et al., 2018; Shayanfar and Jouyban, 2013; Shi et al., 2019). The ratio of API to co-former can be relatively high which helps in the formation of high API dosage tablets (Jensen et al., 2016b; Wang et al., 2019). COAM systems are similar to co-crystals with them both containing two components, usually with one API and one co-former (Karimi-Jafari et al., 2018). The difference between co-crystals and COAM systems is that co-crystals are based on a repeating three dimensional crystal lattice whereas COAM systems have no repeating units and an amorphous structure (Newman et al., 2018). The physical stability of COAM systems is usually higher than that of pure amorphous materials and COAM systems often have improved dissolution characteristics compared to pure amorphous APIs (Löbmann et al., 2013a; Löbmann et al., 2012b). COAM systems are stabilised, for example, by the formation of hydrogen bonds, π-π stacking and ionic bonds between the two compounds, as shown by infrared spectroscopy (Löbmann et al., 2012a; Löbmann et al., 2013b). Methods to produce COAM systems include co-melting, solvent evaporation and mechanochemistry (Chavan et al., 2016). Co-melting involves melting the components followed by rapid cooling to avoid nucleation and recrystallization (Hoppu et al., 2009; Knapik et al., 2015; Teja et al., 2015). A key challenge in co-melting is that some of the APIs or co-formers may thermally degrade if kept at high temperatures for too long (Fan et al., 2019; Goodwin et al., 2018). Solvent evaporation involves dissolving the two components into a solvent or solvent mixture followed by rapidly evaporating the solvent to prevent nucleation and recrystallization (Ahmed Mahmoud Abdelhaleem et al., 2015; Yamamura et al., 2002). However, finding a solvent or solvent mixture which can dissolve both the co-former and the API without one component crystallising prematurely is a challenge (Mishra et al., 2018). Mechanochemistry involves using mechanical stress to reduce crystallinity and induce intimate mixing (Chieng et al., 2009; Hu et al., 2014). The conventional method used for mechanochemistry is milling. A low temperature is preferred during milling to promote the formation of an amorphous

material by keeping the mixture below the glass transition temperature of the amorphous system (Blaabjerg et al., 2017).

The possible co-formers used to form COAM systems are numerous but there is no clear method of predicting whether a certain co-former will form a COAM system with a specific API. Mizoguchi et al. (2019) linked the formation of COAM systems to the mixing enthalpy and the difference in lipophilicity ($\Delta log\ P$). This work used COSMOquick, a computational program which uses the Conductor like Screening Model for Real Solvents (COSMO-RS) method to derive charge density surfaces which describe each molecule and can be used to calculate interaction energies with other components (Klamt, 2018). COSMOquick can be used to screen for potential co-crystals and provides values for the Gibbs energy of mixing (ΔG_{mix}), which determines whether mixing between potential co-crystal formers at constant temperature and pressure is spontaneous, as well as the excess enthalpy of mixing, which is the enthalpy released or absorbed upon mixing (Loschen and Klamt, 2015). Ueda et al. (2016) performed a multivariate analysis of physiochemical variables of co-formers and concluded that a range of these variables (crystallisation tendency, glass transition temperature and molecular flexibility) contributed to COAM formation; however, this study only used one API (naproxen) and a small number of co-formers (felbinac, flufenamic acid, loxoprofen, ketoprofen, indomethacin, aceclofenac, indoprofen).

Meng-Lund et al. (2018) used a range of molecular descriptors to produce a PLS-DA model to predict the likelihood of success of co-amorphisation between amino acids and an API. The model used a dataset formed from 6 APIs and 20 amino acids from Kasten et al. (2016). The variables used include physical properties, Hückel theory descriptors, subdivided surface areas, atom counts, bond counts, pharmacophore feature descriptors, partial charge

descriptors, surface area, volume and shape descriptors. To test the model, one of the six APIs was left out of the model and used as a validation set. Out of the 20 systems in the validation set, 19 were correctly assigned. The model showed that polar amino acids were less likely to form COAM systems and non-polar side chains were more likely to form COAM systems. However, this model only investigated amino acid co-formers.

The current study aims to develop a method to improve the selection of co-formers to formulate COAM systems. The previously reported COAM screen by Kasten et al. (2016) was used to understand which variables affect the formation of COAM systems. Variables used to describe the systems were obtained using COSMOquick to calculate properties that describe the two-component systems and Pubchem to source physico-chemical variables to describe the co-formers. The 36 variables from COSMOquick and Pubchem were used to develop a partial least squares-discriminant analysis (PLS-DA) prediction method to identify which co-formers are likely to form COAM systems.

2. Materials and Methods

2.1 Materials

Succinic acid was purchased from Avocado Research Chemicals (Heysham, UK). Glycine (GLY) was purchased from BDH Chemicals Limited (Hull, UK). Carvedilol (CAR) was obtained from Cilpa Ltd. (Mumbai, India). L-alanine (ALA), flurbiprofen, furosemide (FUR), L-isoleucine (ILE), L-leucine (LEU), L-lysine (LYS), mebendazole (MEB) and L-tyrosine (TYR) were purchased from Flourochem (Hadfield, UK). Indomethacin (IND) was purchased from Hawkins Pharmaceutical group (Minnesota, USA). Urea was purchased from Lancaster Synthesis (Lancaster, UK). Maleic acid was purchased from M&B Chemicals (London, UK). 3-aminobenzoic acid, 4-aminobenzoic acid, 4-aminosalicylic acid, 5-aminosalicylic acid, L-

arginine (ARG), ascorbic acid, L-asparagine (ASN), L-aspartic acid (ASP), 4,4'-bipyridine, caffeine, catechol, L-cysteine (CYS), 2,4 dihydroxybenzoic acid, 3,5-dihydroxybenzoic acid, fumaric acid, gallic acid, L-glutamine (GLN), L-glutamic acid (GLU), glycolic acid, L-histidine (HIS), imidazole, isonicotinamide, ketoprofen, L-methionine (MET), nicotinamide, oxalic acid, L-phenylalanine (PHE), phenazine, piperazine, piracetam, L-proline (PRO), pyrogallol, salicylic acid, L-serine (SER), tartaric acid, theophylline, L-threonine (THR), L-tryptophan (TRP) and L-valine (VAL) were purchased from Sigma Aldrich (Missouri, USA).

2.2 Mebendazole co-former screening

Ball milling was used to screen for potential COAM systems. A 1:1 molar ratio of API and co-former (total 100 mg), was placed into a 5 mL milling jar and premixed at a frequency of 30 Hz for 5 minutes without a mixing ball to homogenize the material. A stainless-steel ball with a diameter of 5 mm was added and the mixture was milled at 30 Hz for 60 min. The milling time of 60 min was selected due to it matching the original study the model was produced from (Kasten et al., 2016). Milling was performed using a Mixer mill MM200, vibrational ball mill, from Retsch GmbH & Co. (Haan, Germany). The mixtures were analysed by XRPD to assess crystallinity (see below).

2.3 Film casting mebendazole – gallic acid

A 1:1 molar ratio of mebendazole to gallic acid (63.5 mg: 36.5 mg), was dissolved in a minimum amount of formic acid (approx. 10 mL). The solution was cast onto a petri dish and the formic acid was left to evaporate. Once the mixture was dry, it was analysed by XRPD (see below).

2.4 X-ray powder diffraction (XRPD)

XRPD measurements were performed using a Bruker D8 X-ray diffractometer (Billerica, Massachusetts) with CuKα radiation (1.54187 Å), and acceleration voltage and current of 40 kV and 40 mA, respectively. The samples were scanned in reflectance mode between 2° and 35° 2θ with a scan rate of 0.067335° 2θ/s and a step size of 0.026°.

2.5 COSMOquick calculations

COSMOquick version 1.7 (COSMOlogic GmbH & Co. KG, Leverkusen, Germany) was used to calculate the Gibbs energy of mixing (ΔG_{mix}), excess enthalpy of mixing (ΔH_{mix}) and excess enthalpy of hydrogen bonding (ΔH_{hb}), of the two-component system. For each component the following variables were calculated and displayed in Table 1: the number of *Rotatable bonds*; *rotbsdmod*, a general molecular flexibility parameter; *M2*, *M3*, *M4*, *M5* and *M6*, the different order sigma moments; the dielectric energy; the molecular COSMO volume; *Macc1*, *Macc2*, *Macc3* and *Macc4*, the different order sigma acceptor moments; *Mdon1*, *Mdon2*, *Mdon3* and *Mdon4*, the different order sigma donor moments; *avratio*, the surface-volume ratio based on COSMO; *ovality*, the ratio of the molecular COSMO area to the area of a sphere with the same volume as the molecule; μ, the pseudo chemical potential of the pure solute; δd, the Hansen parameter for dispersion; δp, the Hansen parameter for permanent dipole-dipole interaction; δh, the Hansen parameter for hydrogen bonding. The difference between the API and co-former values were calculated and used as the variables in the PLS-DA.

Table 1: The definitions of all the variables used to find the PLS-DA model.

Variable	Definition
ΔG_{mix}	Gibbs energy of mixing.

ΔH_{hb}	Excess enthalpy of hydrogen bonding.
ΔH_{mix}	Excess enthalpy of mixing.
Δlog P	The difference between the log of the octanol/water partition coefficient of the API and the co-former
AV. log P	The average value of the log of the octanol/water partition coefficient of the API and the co-former
ΣHBC_{self}	The sum of the difference of hydrogen bond donors to hydrogen bond acceptors for the individual components, for both the API and co-former. To represent the hydrogen bonding present in the individual components.
$\Sigma HBC_{API-COF}$	The sum of the difference of hydrogen bond donors to hydrogen bond acceptors for the mixed components, for both the API and co-former. To represent the hydrogen bonding between the two components.
AV. TM	The average melting point of the two components.
Δ TM	The difference of the melting point of the co-former and the API.
AV. MW	The average molecular weight of the API and the co-former.
Δ MW	The difference of the molecular weights of the API and the co-former.
AV. TPSA	The average topological polar surface area of the API and the co-former.
ΔTPSA	The difference between the topological polar surface area of the API and the co-former.
Δrotatable bonds	The difference between the number of rotatable bonds of the co-former and the API.
Δrotbsdmod	The difference between the general molecular flexibility parameter of the co-former and the API.
ΔM2	The difference between the second order sigma moments of the co-former and the API.
ΔM3	The difference between the third order sigma moments of the co-former and the API.
ΔM4	The difference between the fourth order sigma moments of the co-former and the API.
ΔM5	The difference between the fifth order sigma moments of the co-former and the API.

ΔM6	The difference between the sixth order sigma moments of the co-former and the API.
ΔDielectric energy	The difference between the number of rotatable bonds of the co-former and the API.
Δvolume	The difference between the dielectric energy of the co-former and the API.
ΔMacc1	The difference between the first order sigma acceptor moments of the co-former and the API.
ΔMacc2	The difference between the second order sigma acceptor moments of the co-former and the API.
ΔMacc3	The difference between the third order sigma acceptor moments of the co-former and the API.
ΔMacc4	The difference between the fourth order sigma acceptor moments of the co-former and the API.
ΔMdon1	The difference between the first order sigma donor moments of the co-former and the API.
ΔMdon2	The difference between the second order sigma donor moments of the co-former and the API.
ΔMdon3	The difference between the third order sigma donor moments of the co-former and the API.
ΔMdon4	The difference between the fourth order sigma donor moments of the co-former and the API.
Δavratio	The difference between the surface-volume ratio based on COSMO of the co-former and the API.
Δovality	The difference between co-former and the API of the ratio of the molecular COSMO area to the area of a sphere with the same volume as the molecule.
Δμ	The difference between the pseudo chemical potential of the pure solute of the API and the co-former.
Δ(δd)	The difference between the Hansen parameter for dispersion in $MPa^{0.5}$ of the API and the co-former.
Δ(δp)	The difference between the Hansen parameter for permanent dipole-dipole interactions in $MPa^{0.5}$ of the API and the co-former.

Δ(δh)	The difference between the Hansen parameter for hydrogen bonding in $MPa^{0.5}$ of the API and the co-former.

2. 6 Partial least squares – discriminant analysis

Partial least squares – discriminant analysis (PLS-DA) was performed using SIMCA V.16 (Umetrics, Umeå, Sweden) to plot 36 variables for each combination of API and co-former (Brereton and Lloyd, 2014; Sadeghi-Bazargani et al., 2010). The 36 different variables plotted were ΔG_{mix}, ΔH_{mix}, ΔH_{hb}, *Δlog P*, *AV.log P*, ΣHBC_{self}, $\Sigma HBC_{API\text{-}COF}$, *AV.TM*, *ΔTM, AV.MW, ΔMW, AV.TPSA, ΔTPSA, ΔRotatable bonds, Δrotbsdmod, ΔM2, ΔM3, ΔM4, ΔM5, ΔM6, ΔDielectric energy, Δvolume, ΔMacc1, ΔMacc2, ΔMacc3, ΔMacc4, ΔMdon1, ΔMdon2, ΔMdon3, ΔMdon4, Δavratio, Δovality, Δμ, Δ(δd), Δ(δp) and Δ(δh)* (Table 1). The data was scaled using unit variance. Each system was assigned as either COAM or not COAM (any crystalline material present) as determined by Kasten et al. (2016) based on analysing the mixture by XRPD after ball milling for 60 minutes. The PLS-DA was fitted using two latent variables and all 36 variables. The quality of the model was assessed using an internal cross-validation procedure which involved leaving one out using seven cross-validation groups.

The prediction ability of the model was assessed by checking the predicted values of COAM formation of the 120 API-amino acid dataset and comparing the values with the experimental results. The prediction gives a predicted numerical value with a value closer to one being COAM and a value closer to zero being not COAM. The prediction of the model was also assessed by using a dataset of 29 co-formers paired with mebendazole. The predicted values for the mebendazole-co-former dataset were compared with the experimental values to determine the prediction ability.

Variable selection was used to reduce the number of variables from 36 to 7 based on optimising the number of correctly predicted samples for the API-amino acid dataset. The final model was produced in JMP Pro 15 to view the equation used to assign the COAM value (Equation 1) (JMP, Version Pro 15. 1989-2020).

$$\begin{aligned}\text{Predicted COAM value} &= (-0.123 \times \Delta H_{hb}) + (-0.136 \times \Delta H_{mix}) + \left(-0.00350 \times \Sigma HBC_{self}\right) \\ &+ (0.00297 \times AV.MW) + (-0.00176 \times \Delta TPSA) + (0.0105 \times \Delta\mu) \\ &+ \left(-0.0441 \times \Delta(\delta h)\right) + (-0.204)\end{aligned}$$

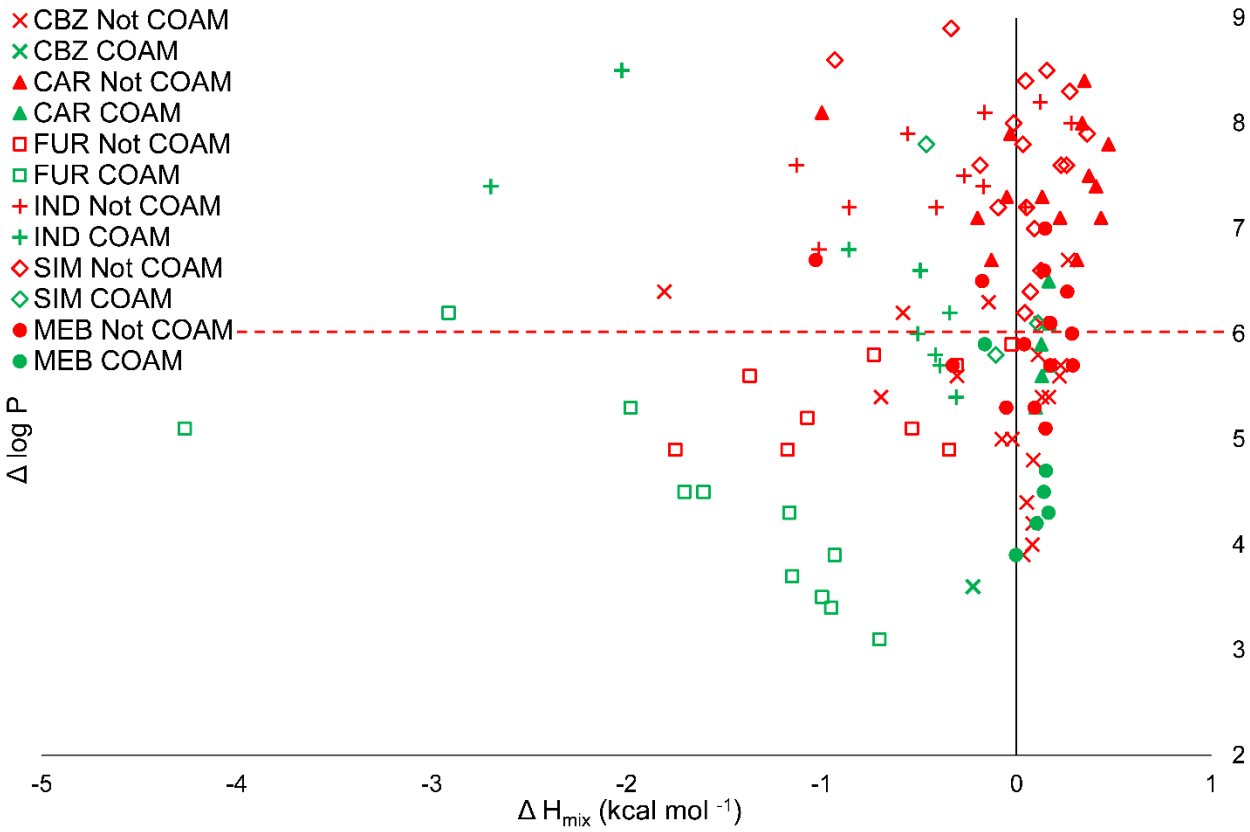
Equation 1: The equation to describe the relation of the seven key variables to the predicted COAM value. All numbers have been rounded to 3 significant figures. A value closer to one indicates the system should be COAM and a value closer to zero indicates it should not be COAM.

3. Results and Discussion

3.1 Correlation of ΔH_{mix} and $\Delta log\ P$ with co-amorphisation

The COAM systems used in this screen were experimentally identified by Kasten et al. (2016) and the responses listed in supplementary materials Table S1 indicate which systems formed COAM materials after 60 min of ball milling. The APIs used were carvedilol (CAR), furosemide (FUR), indomethacin (IND) simvastatin (SIM), carbamazepine (CBZ) and mebendazole (MEB). Previous research on theoretical descriptors for the prediction of the formation of a COAM system identified two indicators (ΔH_{mix} and $\Delta log\ P$) using a combination of APIs with other APIs or sugars to screen for COAM systems using differential scanning calorimetry (Mizoguchi et al., 2019). The ΔH_{mix} was calculated using

COSMOquick and the $\Delta log\ P$ was sourced from Pubchem; it was found that COAM systems form with a $\Delta log\ P$ below 6 and a negative ΔH_{mix} and a clear divide between the COAM systems and the crystalline systems was observed. When the ΔH_{mix} and $\Delta log\ P$ for the API/amino acids systems tested by Kasten et al. (2016) were plotted against each other the same clear divide was not evident (Figure 1). The data indicates that COAM materials tend to form in systems with a lower value of $\Delta log\ P$ and a negative ΔH_{mix}. However, many combinations break these trends; a few COAM systems form with a $\Delta log\ P$ above 6 and many systems with a $\Delta log\ P$ below 6 remain crystalline. Furthermore, COAM systems form with positive values of ΔH_{mix}. To further assess the prediction ability for COAM formation of the two variables a range of 29 different co-formers (supplementary materials Table S2) were paired with mebendazole and analysed using the two variables. The 29 different co-formers were then ball milled with mebendazole to determine whether they formed COAM mixtures and the results were compared with the predicted trends. Figure 2 shows that all the systems including COAM and not COAM have a $\Delta log\ P$ below 6 suggesting $\Delta log\ P$ is not a good predictor of COAM material formation. Figure 2 also shows that the majority of the 29 systems have negative values of ΔH_{mix} but not all the systems are COAM and there is no clear divide between COAM and not COAM systems. Using these two variables to predict the formation of co-amorphous API-co-former systems was insufficient suggesting that more variables were required to predict the propensity to form COAM systems.



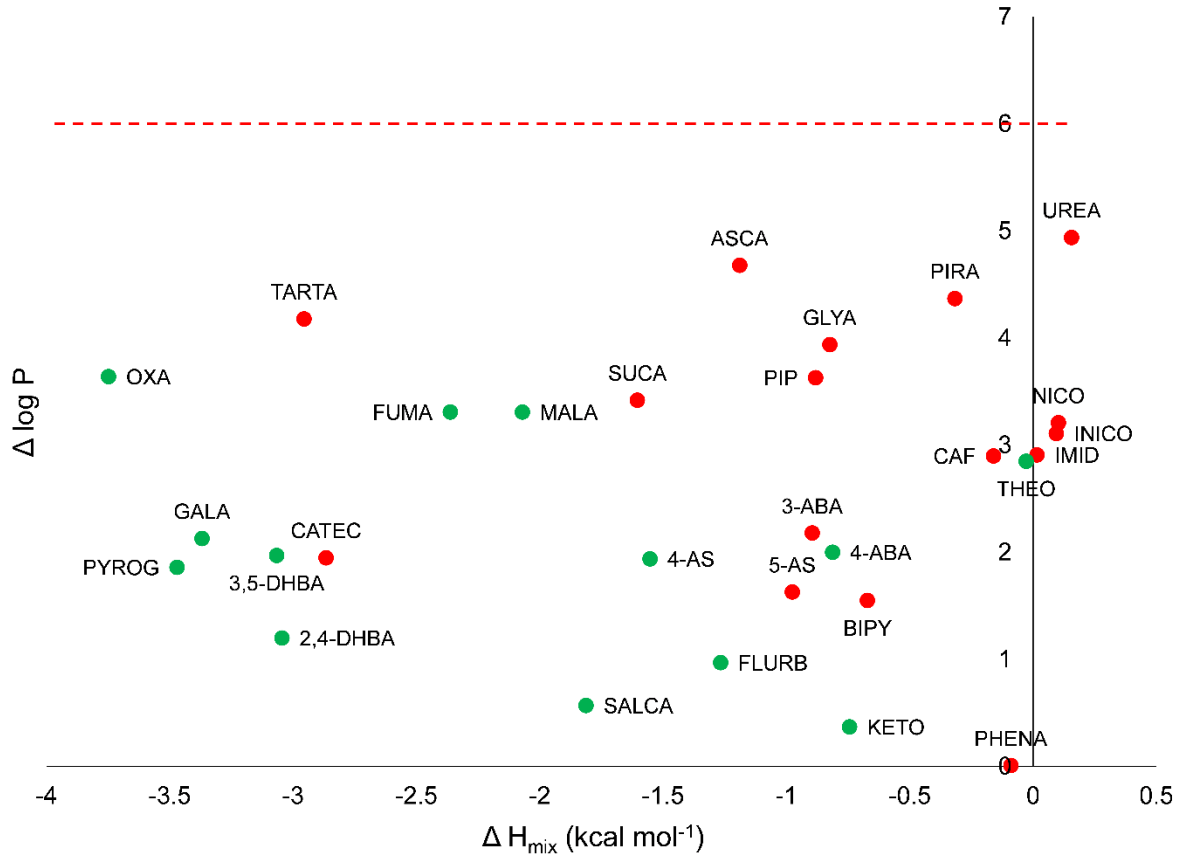
235

236 Figure 1. Relationship between the formation of COAM systems from Kasten et al. (2016),

237 ΔH_{mix} and Δlog P. Green markers indicate COAM systems were formed and red markers

238 indicate not COAM systems. The red dotted line is the expected boundary line between

239 COAM and not COAM systems (Mizoguchi et al., 2019).



241

242 Figure 2. Relationship between the formation of COAM systems of mebendazole with 29 co-
243 formers, ΔH_{mix} and Δlog P. Green markers indicate COAM systems were formed and red
244 markers indicate not COAM systems were formed. The red dotted line is the expected
245 boundary line between COAM and not COAM systems based on previous research by
246 Mizoguchi et al. (2019). Abbreviations of the coformers are as follows: 2,4-dihydroxybenzoic
247 acid (2,4-DHBA), 3,5-dihydroxybenzoic acid (3,5-DHBA), 3-aminobenzoic acid (3-ABA),
248 4,4'-bipryidine (BIPY), 4-aminobenzoic (4-ABA), 4-aminosalicylic acid (4-AS), 5-
249 aminosalicylic acid (5-AS), ascorbic acid (ASCA), caffeine (CAF), catechol (CATEC),
250 flurbiprofen (FLURB), fumaric acid (FUMA), gallic acid (GALA), glycolic acid (GLYA),
251 imidazole (IMID), isonicotinamide (INICO), ketoprofen (KETO), maleic acid (MALA),

nicotinamide (NICO), oxalic acid (OXA), phenazine (PHENA), piperazine (PIP), piracetam (PIRA), pyrogallol (PYROG), salicylic acid (SALCA), succinic acid (SUCA), tartaric acid (TARTA), theophylline (THEO), and urea (UREA).

3.2 PLS-DA

To improve the prediction of COAM systems, 34 additional variables were selected to describe the properties and interactions of the two components and combined with ΔH_{mix} and $\Delta log\ P$. These 36 variables (Table 1) were used to produce a PLS-DA model to understand which variables affect COAM system formation. Variable selection was then used to reduce the initial 36 variables to seven key variables. Variable selection was performed by removing variables one after the other and checking the effect on the prediction ability of the model for the API amino acid data set; if the variable had no effect it was removed and if the prediction ability was reduced it was retained. The variables selected describe differences between the API and co-former allowing the model to be applied to systems where the API and co-former cannot be easily defined, such as systems formed from two APIs. The final PLS-DA model includes the seven descriptors (Table 1): ΔH_{hb}, ΔH_{mix}, ΣHBC_{self}, $AV.\ MW$, $\Delta TPSA$, $\Delta\mu$ and $\Delta(\delta h)$. The goodness of fit is R^2Y = 33.0 %, R^2X = 47.8 % and the goodness of prediction is Q^2 = 29.0 % based on two latent variables. Latent variables are variables which cannot be measured and are inferred from mathematical models.

3.3 Model

The score scatter plot of the PLS-DA model for the API amino acid systems (Figure 3) shows a division between COAM and not COAM systems with COAM systems appearing more in the top right quadrant. The dotted line in Figure 3 shows the predicted separation for visualization purposes between the COAM and not COAM systems. The not COAM systems

occur on the left of the plot and mainly in the bottom left quadrant. Equation 1 shows the relationship of each variable to the overall prediction.

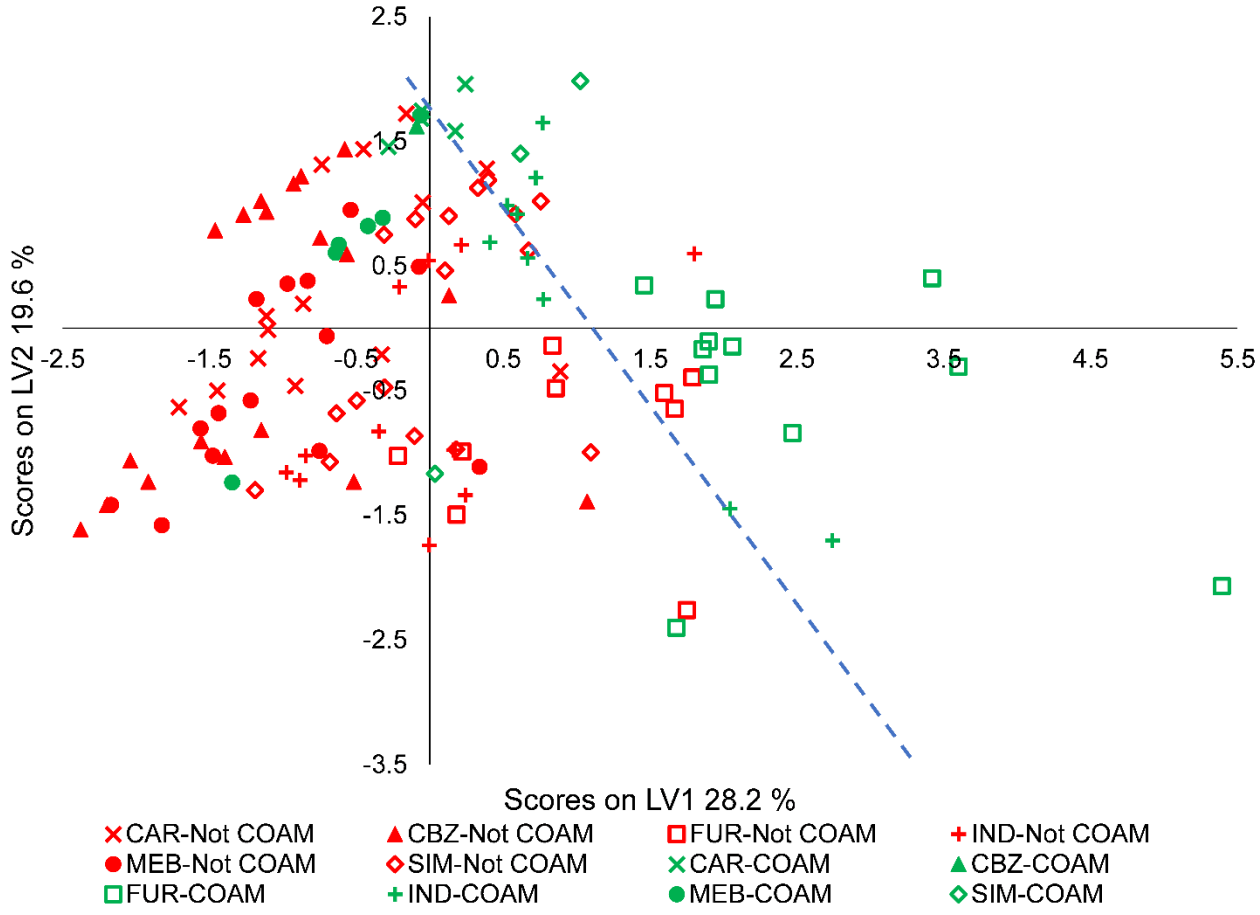
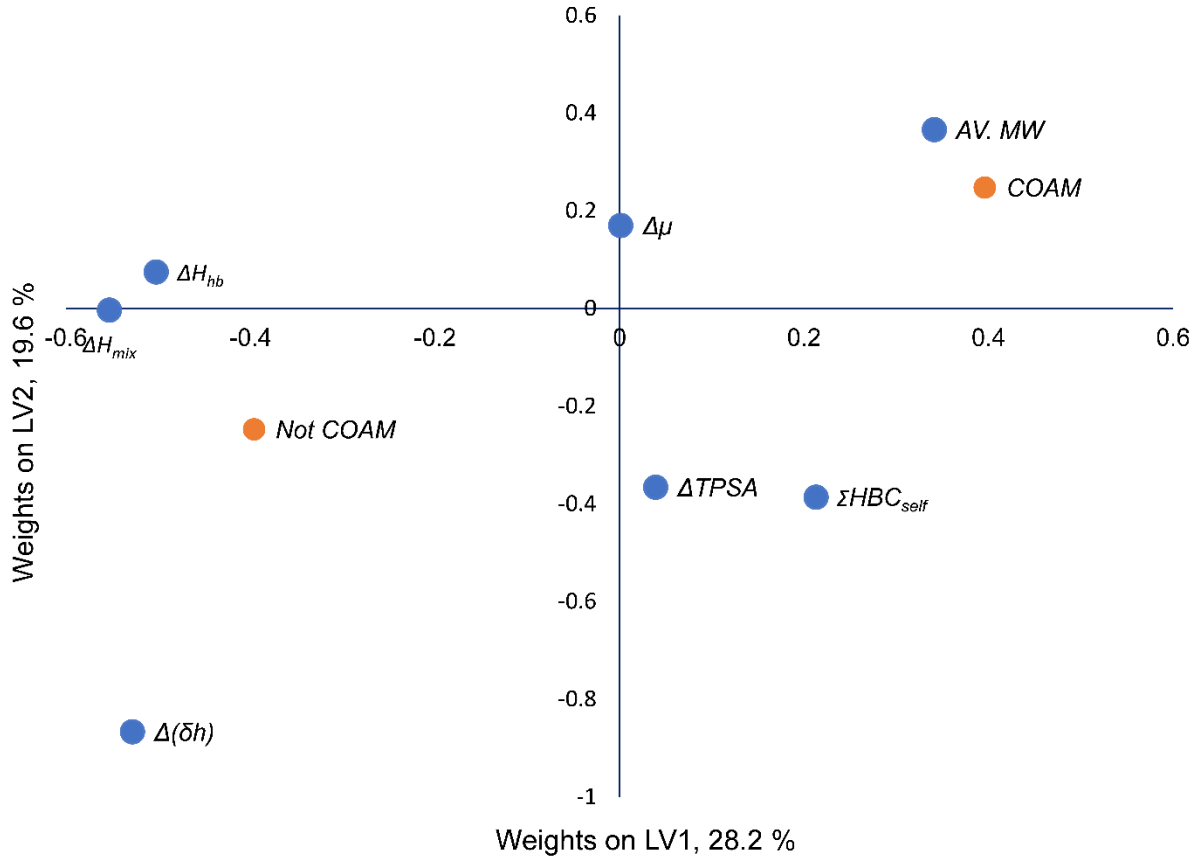


Figure 3. PLS-DA score scatter plot of latent variables (LV) 1 and 2. The red markers indicate not COAM systems and the green markers show COAM systems. The APIs are displayed with different markers with carvedilol (CAR) displayed as an X, carbamazepine (CBZ) as a triangle, furosemide (FUR) as a hollow square, indomethacin (IND) as a cross, mebendazole (MEB) as a circle and simvastatin (SIM) as a hollow diamond. The dashed blue line shows the predicted separation between COAM and Not COAM systems for visualization purposes.

The loading plot (Figure 4) shows how each variable is related to COAM formation. The variables closest to the COAM response are linked to COAM formation and the variables closest to the not COAM response are linked to not COAM formation. The two variables $\Delta TPSA$ and $\Delta\mu$ are located roughly in the middle between the COAM and not COAM point and therefore, do not appear to influence the COAM formation to a strong degree, however, when they are removed the prediction ability of the model is reduced. The variables related to COAM formation, therefore, appear to be a relatively large value of *AV*. *MW* and ΣHBC_{self}, and a relatively small or negative value of ΔH_{mix}, ΔH_{hb}, and $\Delta(\delta h)$. A large *AV*.*MW* seems to correlate with COAM formation possibly due to slower diffusion which would inhibit recrystallization. A large value of ΣHBC_{self} correlates with COAM formation, which is expected due to molecules that do not have a similar number of hydrogen bond donor atoms and hydrogen bond acceptor atoms are unlikely to form as strong crystal structures and may be more likely to interact with the other component (Corpinot and Bučar, 2019). A negative value of ΔH_{mix} favours COAM formation, as expected since negative values indicate that the mixed system has a lower free energy state due to stronger attractive forces between the mixed molecules compared to the individual component interaction. A negative value ΔH_{hb} also favours COAM formation which is due to stronger hydrogen bonding between the mixed molecules when compared with the individual components. A small $\Delta(\delta h)$ seems to favour COAM formation suggesting molecules with similar hydrogen-bonding potential are more likely to interact and stabilise a COAM system.



305

306 Figure 4. PLS-DA loading weights scatter plot of the latent variables (LV) 1 and 2. The
307 responses are shown with orange circles and the variables with blue circles. The responses
308 show how the two groups are related to the variables.

309 The score plot (Figure 3) shows the two clusters of COAM and not COAM samples overlap
310 to some degree. Overall, the misclassification table (which shows if the prediction matches
311 the experimental result) (Table 2) of the 120 API – amino acid dataset shows that 81 % of the
312 data points are correctly placed, suggesting the PLS-DA model is successful at modelling the
313 amino acid data. Out of the 23 misplaced systems, 18 are close to the separation line and five
314 are very far from the separation line, these five systems are MEB with LYS, LEU and ILE,
315 SIM with LYS, and IND with HIS. The MEB with LYS, LEU and ILE and SIM with LYS
316 systems were shown by Kasten et al. (2019) to have a low stability and underwent

crystallisation within a few weeks suggesting the model helps identify stable COAM systems. The fifth system furthest from the separation line was IND with HIS which was not COAM by milling, however a study by Jensen et al. (2016a) showed IND with HIS system was co-amorphous when spray dried, this suggest the model could be valid for other co-amorphous production methods.

Table 2. Misclassification table showing the percentage of correctly assigned observations of the 120 API-amino acid combinations. Fisher's probability of 4.7×10^{-8}.

Model	Members	Correct	Not COAM	COAM
Not COAM	84	90.48%	76	8
COAM	36	58.33%	15	21
Total	120	80.83%	91	29

3.5 Prediction of co-amorphous formation by mebendazole with 29 co-formers

To test the applicability of the PLS-DA model (Figure 3, Figure 4) to other non-amino acid systems, a new dataset of 29 different co-formers with mebendazole was used. Mebendazole was selected due to it forming a range of both of COAM and not COAM systems with the amino acids; therefore, it was expected to form a range of both co-amorphous and not COAM systems with other co-formers. The 29 co-formers were selected on the basis of being small molecules capable of forming a range of different hydrogen-bonding motifs. The model was applied to predict the classification of the mebendazole co-former mixtures and the prediction was compared to experimental data. The misclassification table (Table 3) shows that overall, 86 % of the samples were predicted correctly and only four of 29 mixtures were predicted incorrectly. The score plot of the predicted scores (Figure 5) shows a clear divide between

systems that were COAM and systems that were not COAM, with only a slight overlap of the two clusters.

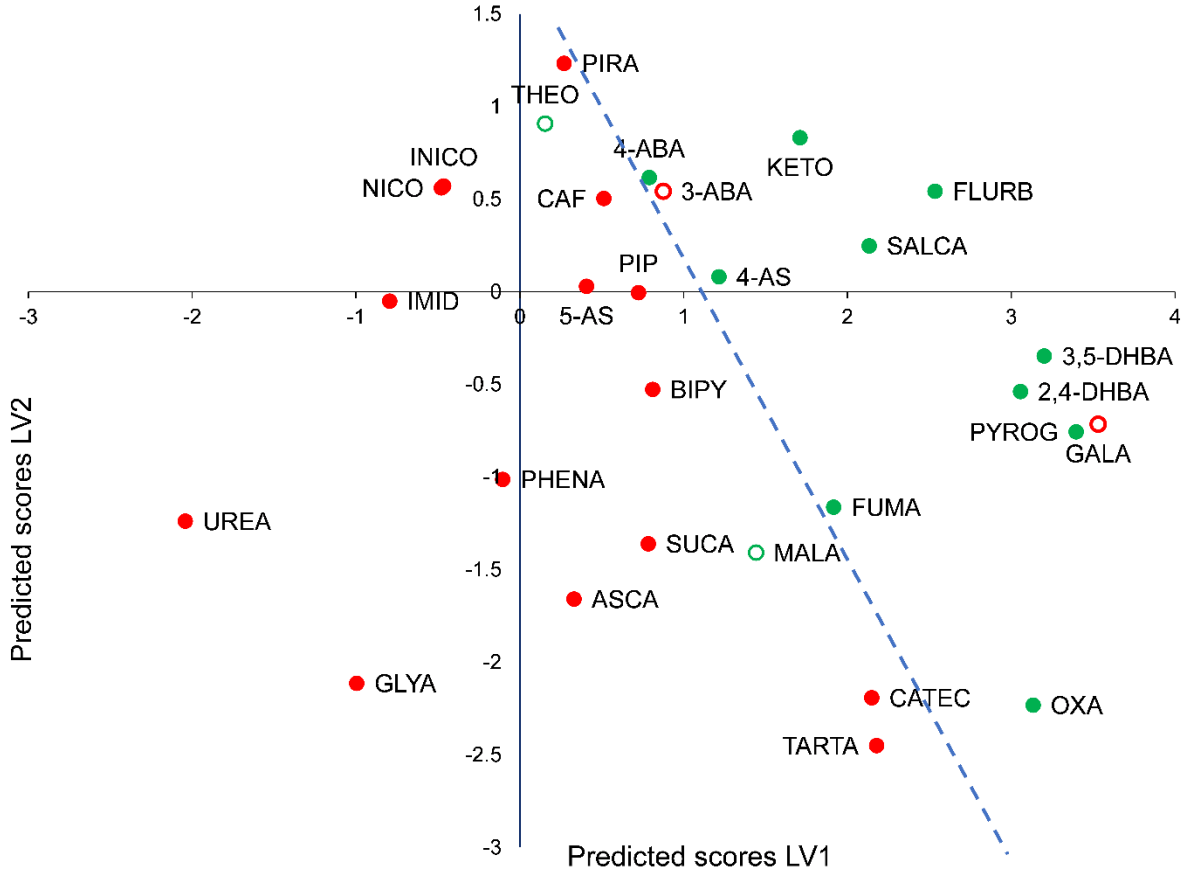


Figure 5. Score scatter plot of the predicted scores for the mebendazole-co-former combinations. COAM samples are shown in green, not COAM samples are shown in red. The hollow circles indicate samples which have been predicted incorrectly. The blue dashed line shows the predicted separation line for visualization purposes.

Table 3. Misclassification table showing the percentage of correctly assigned observation of the 29 MEB-co-former combinations. Fisher's probability of 1.8×10^{-4}.

Model	Members	Correct	Not COAM	COAM

Not COAM	17	88.24%	15	2
COAM	12	83.33%	2	10
Total	29	86.21%	17	12

The four samples that were predicted incorrectly were MEB combinations with theophylline, 3-aminobenzoic acid, maleic acid and gallic acid, with predicted COAM values of 0.43, 0.52, 0.40 and 0.86, respectively. The COAM values indicate how close the prediction is to assigning the system as COAM or not COAM with a value above 0.5 indicating COAM and a value below 0.5 indicating not COAM. Three of the samples had COAM values close to the cross over point at 0.5, suggesting they were close to being predicted correctly and may have been misplaced by one or two variables having extreme values. Theophylline appears to be incorrectly predicted due to the system having a relatively high value of ΔH_{mix} and ΔH_{hb} compared to the other systems. 3-aminobenzoic acid was misplaced due to a relatively small/negative ΔH_{mix} and ΔH_{hb} and a small $\Delta(\delta h)$. Maleic acid is misplaced due to a small MW. The gallic acid system is the furthest away from the crossover line between COAM and not COAM systems, suggesting it should be COAM. The mebendazole gallic acid system was investigated using film casting which resulted in a COAM system, this suggests the model is not limited to system produced by only ball milling. Film casting was selected because it involves a thermodynamic pathway with the initial solution containing no crystalline material compared to ball milling which is a kinetic pathway involving the disruption of the crystal lattice (Karagianni et al., 2018). Therefore, film casting is likely to help the formation of a co-amorphous system if the initial crystalline material is too stable to be broken down by ball milling. With the mebendazole gallic acid system now being classed as COAM the misclassification table improves and the correct prediction percentage is now 90 % (Table 4). The model now shows an even clearer divide between the two clusters with

only a few outliers which are close to the cross over line. Film casting was not used to test other systems due to other co-amorphous formation methods usually producing a similar result (Karmwar et al., 2011; Lim et al., 2016).

Table 4: Misclassification table showing the percentage of correctly assigned observations of the 29 MEB-co-former combinations after gallic acid was confirmed as being COAM by film casting. Fisher's probability of 2.4×10^{-5}.

Model	Members	Correct	Not COAM	COAM
Not COAM	16	93.75%	15	1
COAM	13	84.62%	2	11
Total	29	89.66%	17	12

4. Conclusion

Known COAM systems formed with APIs and amino acid co-formers were analysed to identify properties of the co-former that correlate with COAM material formation (Kasten et al., 2016). A range of 36 variables was used to describe the properties of the API-amino acid systems and a multivariate PLS-DA was used to create a prediction model. The initial 36 variables were reduced to seven variables including ΔH_{hb}, ΔH_{mix}, ΣHBC_{self}, $AV.\ MW$, $\Delta TPSA$, $\Delta\mu$ and $\Delta(\delta h)$. The model predicts 81 % of the API-amino acid systems correctly. The model was tested using a dataset of mebendazole with 29 different co-formers and 90 % of the systems were correctly predicted. Overall, the model can predict the potential COAM

formation of a range of co-formers significantly expanding its applicability beyond the relatively limited set of amino acid co-formers.

CRediT authorship contribution statement

Luke I. Chambers: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – Original Draft, Visualization, Project administration. **Holger Grohganz**: Conceptualization, Methodology, Writing – Review & Editing, Supervision. **Henrik Palmelund**: Validation, Investigation, Writing – Review & Editing. **Korbinian Löbmann**: Conceptualization, Writing – Review & Editing, Supervision. **Thomas Rades:** Conceptualization, Writing – Review & Editing, Supervision. **Osama M. Musa**: Supervision, Funding acquisition. **Jonathan W. Steed**: Conceptualization, Writing – Review & Editing, Supervision.

Conflicts of interest

The author declares no conflicts of interest.

Acknowledgements

We acknowledge the Engineering and Physical Sciences Research Council for funding, through the Soft Matter and Functional Interfaces Centre for Doctoral Training and Ashland LLC for studentship funding.

References

JMP, Version Pro 15. SAS Institute Inc., Cary, NC, 1989-2020.

Ahmed Mahmoud Abdelhaleem, A., Adel Ahmed, A., Ibrahim Abdullah, M., 2015. Clozapine-carboxylic acid plasticized co-amorphous dispersions: Preparation, characterization and solution stability evaluation. Acta Pharm. 65, 133-146.

Baghel, S., Cathcart, H., O'Reilly, N.J., 2016. Polymeric Amorphous Solid Dispersions: A Review of Amorphization, Crystallization, Stabilization, Solid-State Characterization, and Aqueous Solubilization of Biopharmaceutical Classification System Class II Drugs. J. Pharm. Sci. 105, 2527-2544.

Berry, D.J., Steed, J.W., 2017. Pharmaceutical cocrystals, salts and multicomponent systems; intermolecular interactions and property based design. Adv. Drug. Deliv. Rev. 117, 3-24.

Blaabjerg, L.I., Lindenberg, E., Rades, T., Grohganz, H., Lobmann, K., 2017. Influence of preparation pathway on the glass forming ability. Int. J. Pharm. 521, 232-238.

Brereton, R.G., Lloyd, G.R., 2014. Partial least squares discriminant analysis: taking the magic away. J. Chemom. 28, 213-225.

Chavan, R.B., Rathi, S., Jyothi, V.G.S.S., Shastri, N.R., 2019. Cellulose based polymers in development of amorphous solid dispersions. Asian J. Pharm. Sci. 14, 248-264.

Chavan, R.B., Thipparaboina, R., Kumar, D., Shastri, N.R., 2016. Co amorphous systems: A product development perspective. Int. J. Pharm. 515, 403-415.

Chieng, N., Aaltonen, J., Saville, D., Rades, T., 2009. Physical characterization and stability of amorphous indomethacin and ranitidine hydrochloride binary systems prepared by mechanical activation. Eur. J. Pharm. Biopharm. 71, 47-54.

Corpinot, M.K., Bučar, D.-K., 2019. A Practical Guide to the Design of Molecular Crystals. Cryst. Growth Des. 19, 1426-1453.

Di, L., Fish, P.V., Mano, T., 2012. Bridging solubility between drug discovery and development. Drug. Discov. Today 17, 486-495.

Fan, W., Zhu, W., Zhang, X., Xu, Y., Di, L., 2019. Application of the combination of ball-milling and hot-melt extrusion in the development of an amorphous solid dispersion of a poorly water-soluble drug with high melting point. RSC Adv. 9, 22263-22273.

Frank, D.S., Matzger, A.J., 2018. Probing the Interplay between Amorphous Solid Dispersion Stability and Polymer Functionality. Mol. Pharm. 15, 2714-2720.

Gao, Y., Liao, J., Qi, X., Zhang, J., 2013. Coamorphous repaglinide–saccharin with enhanced dissolution. Int. J. Pharm. 450, 290-295.

Goodwin, M.J., Musa, O.M., Berry, D.J., Steed, J.W., 2018. Small-Molecule Povidone Analogues in Coamorphous Pharmaceutical Phases. Cryst. Growth Des. 18, 701-709.

Healy, A.M., Worku, Z.A., Kumar, D., Madi, A.M., 2017. Pharmaceutical solvates, hydrates and amorphous forms: A special emphasis on cocrystals. Adv. Drug Deliv. Rev. 117, 25-46.

435 Hoppu, P., Hietala, S., Schantz, S., Juppo, A.M., 2009. Rheology and molecular mobility of
436 amorphous blends of citric acid and paracetamol. Eur. J. Pharm. Biopharm. 71, 55-63.

437 Hu, Y., Gniado, K., Erxleben, A., McArdle, P., 2014. Mechanochemical Reaction of Sulfathiazole
438 with Carboxylic Acids: Formation of a Cocrystal, a Salt, and Coamorphous Solids. Cryst. Growth
439 Des. 14, 803-813.

440 Jensen, K.T., Blaabjerg, L.I., Lenz, E., Bohr, A., Grohganz, H., Kleinebudde, P., Rades, T.,
441 Löbmann, K., 2016a. Preparation and characterization of spray-dried co-amorphous drug–
442 amino acid salts. J. Pharm. Pharmacol. 68, 615-624.

443 Jensen, K.T., Larsen, F.H., Lobmann, K., Rades, T., Grohganz, H., 2016b. Influence of variation
444 in molar ratio on co-amorphous drug-amino acid systems. Eur. J. Pharm. Biopharm. 107, 32-
445 39.

446 Kalepu, S., Nekkanti, V., 2015. Insoluble drug delivery strategies: review of recent advances
447 and business prospects. Acta Pharm. Sin. B. 5, 442-453.

448 Karagianni, A., Kachrimanis, K., Nikolakakis, I., 2018. Co-Amorphous Solid Dispersions for
449 Solubility and Absorption Improvement of Drugs: Composition, Preparation, Characterization
450 and Formulations for Oral Delivery. Pharmaceutics 10, 98.

451 Karimi-Jafari, M., Padrela, L., Walker, G.M., Croker, D.M., 2018. Creating Cocrystals: A Review
452 of Pharmaceutical Cocrystal Preparation Routes and Applications. Cryst. Growth Des. 18,
453 6370-6387.

454 Karmwar, P., Graeser, K., Gordon, K.C., Strachan, C.J., Rades, T., 2011. Investigation of
455 properties and recrystallisation behaviour of amorphous indomethacin samples prepared by
456 different methods. Int. J. Pharm. 417, 94-100.

457 Kasten, G., Grohganz, H., Rades, T., Lobmann, K., 2016. Development of a screening method
458 for co-amorphous formulations of drugs and amino acids. Eur. J. Pharm. Sci. 95, 28-35.

459 Kasten, G., Lobmann, K., Grohganz, H., Rades, T., 2019. Co-former selection for co-amorphous
460 drug-amino acid formulations. Int. J. Pharm. 557, 366-373.

461 Khadka, P., Ro, J., Kim, H., Kim, I., Kim, J.T., Kim, H., Cho, J.M., Yun, G., Lee, J., 2014.
462 Pharmaceutical particle technologies: An approach to improve drug solubility, dissolution and
463 bioavailability. Asian J. Pharm. Sci. 9, 304-316.

464 Khodadadi, S., Meesters, G.M.H., 2018. Amorphous APIs: Improved Release, Preparation,
465 Characterization, in: Merkus, H.G., Meesters, G.M.H., Oostra, W. (Eds.), Particles and
466 Nanoparticles in Pharmaceutical Products: Design, Manufacturing, Behavior and
467 Performance. Springer International Publishing, Cham, Switzerland, pp. 329-346.

468 Kissi, E.O., Grohganz, H., Lobmann, K., Ruggiero, M.T., Zeitler, J.A., Rades, T., 2018. Glass-
469 Transition Temperature of the beta-Relaxation as the Major Predictive Parameter for
470 Recrystallization of Neat Amorphous Drugs. J. Phys. Chem. B 122, 2803-2808.

Klamt, A., 2018. The COSMO and COSMO-RS solvation models. Wiley Interdiscip. Rev. Comput. Mol. Sci. 8, e1338.

Knapik, J., Wojnarowska, Z., Grzybowska, K., Jurkiewicz, K., Tajber, L., Paluch, M., 2015. Molecular Dynamics and Physical Stability of Coamorphous Ezetimib and Indapamide Mixtures. Mol. Pharmaceutics 12, 3610-3619.

Lim, A.W., Löbmann, K., Grohganz, H., Rades, T., Chieng, N., 2016. Investigation of physical properties and stability of indomethacin–cimetidine and naproxen–cimetidine co-amorphous systems prepared by quench cooling, coprecipitation and ball milling. J. Pharm. Pharmacol. 68, 36-45.

Löbmann, K., Grohganz, H., Laitinen, R., Strachan, C., Rades, T., 2013a. Amino acids as co-amorphous stabilizers for poorly water soluble drugs--Part 1: preparation, stability and dissolution enhancement. Eur. J. Pharm. Biopharm. 85, 873-881.

Löbmann, K., Laitinen, R., Grohganz, H., Strachan, C., Rades, T., Gordon, K., 2012a. A theoretical and spectroscopic study of co-amorphous naproxen and indomethacin. Int. J. Pharm. 453, 80-87.

Löbmann, K., Laitinen, R., Strachan, C., Rades, T., Grohganz, H., 2013b. Amino acids as co-amorphous stabilizers for poorly water-soluble drugs--Part 2: molecular interactions. Eur. J. Pharm. Biopharm. 85, 882-888.

Löbmann, K., Strachan, C., Grohganz, H., Rades, T., Korhonen, O., Laitinen, R., 2012b. Co-amorphous simvastatin and glipizide combinations show improved physical stability without evidence of intermolecular interactions. Eur. J. Pharm. Biopharm. 81, 159-169.

Loschen, C., Klamt, A., 2015. Solubility prediction, solvate and cocrystal screening as tools for rational crystal engineering. J. Pharm. Pharmacol. 67, 803-811.

Ma, X., Williams, R.O., 2019. Characterization of amorphous solid dispersions: An update. J. Drug Deliv. Sci. Tec. 50, 113-124.

Marsac, P.J., Konno, H., Rumondor, A.C., Taylor, L.S., 2008. Recrystallization of nifedipine and felodipine from amorphous molecular level solid dispersions containing poly(vinylpyrrolidone) and sorbed water. Pharm. Res. 25, 647-656.

Medarević, D., Djuriš, J., Barmpalexis, P., Kachrimanis, K., Ibrić, S., 2019. Analytical and Computational Methods for the Estimation of Drug-Polymer Solubility and Miscibility in Solid Dispersions Development. Pharmaceutics 11, 372.

Meng-Lund, H., Kasten, G., Jensen, K.T., Poso, A., Pantsar, T., Rades, T., Rantanen, J., Grohganz, H., 2018. The use of molecular descriptors in the development of co-amorphous formulations. Eur. J. Pharm. Sci. 119, 31-38.

Mishra, J., Rades, T., Lobmann, K., Grohganz, H., 2018. Influence of Solvent Composition on the Performance of Spray-Dried Co-Amorphous Formulations. Pharmaceutics 10, 47.

507 Mizoguchi, R., Waraya, H., Hirakura, Y., 2019. Application of Co-Amorphous Technology for
508 Improving the Physicochemical Properties of Amorphous Formulations. Mol. Pharm. 16,
509 2142-2152.

510 Newman, A., Reutzel-Edens, S.M., Zografi, G., 2018. Coamorphous Active Pharmaceutical
511 Ingredient–Small Molecule Mixtures: Considerations in the Choice of Coformers for
512 Enhancing Dissolution and Oral Bioavailability. J. Pharm. Sci. 107, 5-17.

513 Nielsen, L.H., Rades, T., Müllertz, A., 2015. Stabilisation of amorphous furosemide increases
514 the oral drug bioavailability in rats. Int. J. Pharm. 490, 334-340.

515 Rams-Baron, M., Jachowicz, R., Boldyreva, E., Zhou, D., Jamroz, W., Paluch, M., 2018. Physical
516 Instability: A Key Problem of Amorphous Drugs, Amorphous Drugs: Benefits and Challenges.
517 Springer International Publishing, Cham, Switzerland, pp. 107-157.

518 Rumondor, A.C., Marsac, P.J., Stanford, L.A., Taylor, L.S., 2009. Phase behavior of
519 poly(vinylpyrrolidone) containing amorphous solid dispersions in the presence of moisture.
520 Mol. Pharm. 6, 1492-1505.

521 Sadeghi-Bazargani, H., Banani, A., Mohammadi, S., 2010. Using SIMCA statistical software
522 package to apply orthogonal projections to latent structures modeling. 2010 World
523 Automation Congress, 1-9.

524 Savjani, K.T., Gajjar, A.K., Savjani, J.K., 2012. Drug solubility: importance and enhancement
525 techniques. ISRN Pharm. 2012, 195727-195727.

526 Shayanfar, A., Jouyban, A., 2013. Drug–Drug Coamorphous Systems: Characterization and
527 Physicochemical Properties of Coamorphous Atorvastatin with Carvedilol and Glibenclamide.
528 J. Pharm. Innov. 8, 218-228.

529 Shi, Q., Moinuddin, S.M., Cai, T., 2019. Advances in coamorphous drug delivery systems. Acta.
530 Pharm. Sin. B 9, 19-35.

531 Sun, Y., Zhu, L., Wu, T., Cai, T., Gunn, E.M., Yu, L., 2012. Stability of amorphous pharmaceutical
532 solids: crystal growth mechanisms and effect of polymer additives. AAPS J. 14, 380-388.

533 Teja, A., Musmade, P.B., Khade, A.B., Dengale, S.J., 2015. Simultaneous improvement of
534 solubility and permeability by fabricating binary glassy materials of Talinolol with Naringin:
535 Solid state characterization, in-vivo in-situ evaluation. Eur. J. Pharm. Sci. 78, 234-244.

536 Tian, Y., Jones, D.S., Andrews, G.P., 2015. An Investigation into the Role of Polymeric Carriers
537 on Crystal Growth within Amorphous Solid Dispersion Systems. Mol. Pharm. 12, 1180-1192.

538 Ueda, H., Muranushi, N., Sakuma, S., Ida, Y., Endoh, T., Kadota, K., Tozuka, Y., 2016. A Strategy
539 for Co-former Selection to Design Stable Co-amorphous Formations Based on
540 Physicochemical Properties of Non-steroidal Inflammatory Drugs. Pharm. Res. 33, 1018-1029.

Van Den Mooter, G., 2012. The use of amorphous solid dispersions: A formulation strategy to overcome poor solubility and dissolution rate. Drug Discov. Today Technol. 9, e71-e174.

Vasconcelos, T., Marques, S., das Neves, J., Sarmento, B., 2016. Amorphous solid dispersions: Rational selection of a manufacturing process. Adv. Drug Deliv. Rev. 100, 85-101.

Wang, Z., Sun, M., Liu, T., Gao, Z., Ye, Q., Tan, X., Hou, Y., Sun, J., Wang, D., He, Z., 2019. Co-amorphous solid dispersion systems of lacidipine-spironolactone with improved dissolution rate and enhanced physical stability. Asian J. Pharm. Sci. 14, 95-103.

Williams, H.D., Trevaskis, N.L., Charman, S.A., Shanker, R.M., Charman, W.N., Pouton, C.W., Porter, C.J., 2013. Strategies to address low drug solubility in discovery and development. Pharmacol. Rev. 65, 315-499.

Wu, W., Lobmann, K., Schnitzkewitz, J., Knuhtsen, A., Pedersen, D.S., Grohganz, H., Rades, T., 2018. Aspartame as a co-former in co-amorphous systems. Int. J. Pharm. 549, 380-387.

Yamamura, S., Gotoh, H., Sakamoto, Y., Momose, Y., 2002. Physicochemical properties of amorphous salt of cimetidine and diflunisal system. Int. J. Pharm. 241, 213-221.